Exercise 1: Kullback-Leibler Divergence

(a) You want to approximate the binomial distribution with n number of trials and probability p with a Gaussian distribution with mean μ and variance σ^2. To find a suitable distribution you investigate the Kullback-Leibler divergence (KLD) in terms of the parameters $\boldsymbol{\theta} = (\mu, \sigma^2)^\top$.

  (i) Write down the KLD for the given setup.
  (ii) Derive the gradients with respect to $\boldsymbol{\theta}$.
  (iii) Is there an analytic solution for the optimal parameter setting? If yes, derive the corresponding solution. If no, give a short reasoning.
  (iv) Independent of the previous exercise, state a numerical procedure to minimize the KLD.

(b) Sample points according to the true distribution and visualize the KLD for different parameter settings of the Gaussian distribution (including the optimal one if available).

(c) Create a surface plot with axes n and p and colour value equal to the KLD for the optimal normal distribution.

(d) Based on the previous result,

  (i) how can the behaviour for varying p be explained?
  (ii) how can the behaviour for varying n be explained?

Exercise 2: The Convexity of KL Divergence

Let p and q be the PDFs of a pair of absolutely continuous distributions.

(a) Prove that the KL divergence is convex in the pair (p, q), i.e.,

$$D_{KL}(\lambda p_1 + (1-\lambda)p_2 || \lambda q_1 + (1-\lambda)q_2) \leq \lambda D_{KL}(p_1||q_1) + (1-\lambda)D_{KL}(p_2||q_2), \tag{1}$$

where (p_1, q_1) and (p_2, q_2) are two pairs of distributions and $0 \leq \lambda \leq 1$.

Hint: you can use the log sum inequality, namely that $(a_1 + a_2)\log\left(\frac{a_1+a_2}{b_1+b_2}\right) \leq a_1 \log\frac{a_1}{b_1} + a_2 \log\frac{a_2}{b_2}$ holds for $a_1, a_2, b_1, b_2 \geq 0$.